DOCUMENT RESUME

ED 446 100 TM 031 843

AUTHOR Arter, Judith
TITLE Rubrics, Scoring Guides, and Performance Criteria: Classroom Tools for Assessing and Improving Student Learning.
PUB DATE 2000-04-00
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000). Based on the book "Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance" by Judith Arter and Jay McTighe, Corwin Press, in press.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Criteria; Elementary Secondary Education; Learning; *Performance Based Assessment; *Scoring Rubrics; *Student Evaluation

ABSTRACT

Rubrics, scoring guides, and performance criteria help define important outcomes for students. Well crafted rubrics help teachers define learning targets so that they can plan instruction more effectively, be more consistent in scoring student work, and be more systematic in reporting student progress. Rubrics can generally be divided into holistic or analytical trait rubrics, and task-specific or general rubrics. A third dimension that distinguishes rubrics is the number of score points. Rubrics must be of high quality in order to have positive effects in the classroom. A metarubric, a rubric for rubrics, has four traits: content, clarity, practicality, and technical soundness. Attachments discuss building a performance rubric and contain a discussion of the metarubric. A final attachment (figure 5) presents seven strategies for using criteria as a teaching tool. (SLD)

Rubrics, Scoring Guides, and Performance Criteria: Classroom Tools for Assessing and Improving Student Learning

Dr. Judith Arter

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans
April 2000

Rubrics, Scoring Guides, and Performance Criteria: Classroom Tools for Assessing and Improving Student Learning[1]

By Dr. Judith A. Arter, Assessment Training Institute, Portland, Oregon

Rubrics, scoring guides, and performance criteria[2] describe what to look for in products or performances to judge their quality. There are essentially two uses for rubrics in the classroom:

1. To gather information on students in order to plan instruction, track student progress toward important learning targets, and report progress to others.
2. To help students become increasingly proficient on the very performances and products also being assessed. In other words, criteria used to enhance the quality of student performance, not simply evaluate it.

The idea in the former is that rubrics, scoring guides, and performance criteria help define important outcomes for students. Many times, some of the more complex outcomes for students are not very well defined in teachers' minds. What is critical thinking, life-long learning, communication in mathematics, etc. and how will teachers know when students do these things adequately? Well crafted rubrics help teachers define these learning targets so that they can plan instruction more effectively, be more consistent scoring student work, and be more systematic in reporting student progress.

The idea in the latter is that when students know the criteria for quality in advance of their performance, they are provided with clear goals for their work. They don't have to guess about what is most important or how their performance will be judged. Further, students learn these criteria and can use them over and over again, deepening their understanding of quality with time. George Hillocks (1986) stated it well:

> "Scales, criteria, and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them. These treatments are two times more effective than free writing techniques."

There is tantalizing evidence that using criteria in these two ways has an impact on teaching and student achievement (Arter, et al., 1995; Borko et al., 1997; Clarke & Stephens, 1996; Khattri, 1995; Hillocks, 1986; NSDC, 2000; OERI, 1997).

The use of rubrics to enhance student learning is a specific (and classic) example of the more general principle of student involvement in their own assessment. Student involvement can take various forms—developing and using assessments, self-assessment, tracking one's own progress, or communicating about one's progress and success (Stiggins, 1999; Black & Wiliam, 1998). In addition to the specific evidence cited above that student use of rubrics can improve

---

[1] Paper presented at AERA, New Orleans, April 24, 2000, session 1.21—New Paradigms in Assessment Policy and Practice. Based on the book: Judy Arter and Jay McTighe (in press). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*, Corwin Press. Part of a book series edited by Tom Guskey and Bob Marzano.

[2] For convenience in this paper, we'll use the terms rubrics, scoring guides, and performance criteria to mean the same thing—guidelines, rules, or principles by which student responses, products, or performances are judged.

achievement, there is more general evidence that improved classroom assessment, including better quality teacher-made assessments, better communication about achievement, and student-involvement in their own assessment can improve student achievement (Black & Wiliam, 1998; Crooks, 1988). In fact, Black & Wiliam report dramatic impacts of improved classroom assessment on student achievement.

However, rubrics aren't magic; not any old rubric, used in any old way, will necessarily have the effects noted above. There are many types of rubrics, scoring guides, and performance criteria and teachers need to know what they are, when to use them, and how to develop them. Teachers need to know what kind of learning targets are best assessed using a rubric (and which can be adequately assessed with other assessment methods), and the features of rubrics that lend themselves to student use.

Further, teachers need to understand the characteristics of quality rubrics. If a rubric doesn't include the features of work that really define quality, teachers will teach to the wrong target and students will learn to the wrong target. If criteria are not clearly stated in a rubric, they will not be much good in illuminating the nature of quality. And, finally, teachers need to have experience using rubrics as instructional tools in order to maximize their benefits.

The following sections expand on these points.

Types of Rubrics

The major dimensions that seen to distinguish rubrics are shown in Figures 1 and 2—holistic or analytical trait rubrics (Figure 1) and task-specific or general rubrics (Figure 2). Figures 1 and 2 present definitions, examples, and a discussion of when to use each type. A third dimension, not illustrated in these tables, that distinguishes rubrics is the number of score points. We've seen rubrics having anywhere from two (essentially a checklist) to 11 score points (e.g., developmental continuums that describe how student competence develops over time, say in reading or writing). The attached examples have scales ranging from four to six points.

Developing Rubrics

Figure 3 presents a summary of how to develop general rubrics for complex student products and performances. It is important to remember that not all rubrics need to be developed like this, nor will all rubrics be this complex (see Figures 1 and 2). Rather, this process should be reserved for those situations in which the quality criteria for an important learning target are somewhat fuzzy and could benefit from this approach—e.g., critical thinking, writing, and mathematics problem solving, reading fluency, art critique, etc.

Metarubric

Rubrics have to be of high quality in order to have positive effects in the classroom. Figure 4 presents a rubric for rubrics—a metarubric[3]. This metarubric has four traits—content, clarity, practicality, and technical soundness.

[3] This is an analytical trait, general rubric with a five-point scale. This type of rubric was chosen so that teachers could learn the traits of good rubrics in general; so they could generalize quality considerations from one rubric to the next. Task-specific metarubrics—a separate metarubric that only describes features of quality for a particular use, say writing or critical thinking—could not be generalized in this fashion. Thus, we are, ourselves, trying to use the metarubric as a tool to help teachers learn about quality rubrics. Of course the metarubric has to adhere to its

Teaching Criteria to Students

Although having high quality rubrics for use by teachers to assess students and make instructional decisions is undeniably important, the final piece of the puzzle is in place when teachers use rubrics to help students learn to self-assess. The heart of academic competence is self-assessment—knowing what to look for in one's own work to decide what could be improved, and then knowing how to improve it.

But, there is no reason to think that teachers will automatically know how to use criteria with students to teach them to self-assess well. Figure 5 presents seven strategies for teaching students how to self-assess using rubrics. But, please understand that there are some conditions that need to be in place for these strategies to work. For example:

- In order to be effective, the rubrics being taught to students have to be of high quality as defined in Table 4.
- A current procedure is to ask students to develop criteria. This is tricky because students, like anyone else, can develop bad ones. The trick from the teacher's point of view is to guide students into discovering the "right" criteria. To do this, teachers need to have a sense of what the "right" criteria are (Metarubric trait of content)[4].
- It is easy to "over-rubric." Teachers should not do this for everything. They need to pick and choose the products or skills that would most benefit from this emphasis.

Future Research

Areas that could benefit from continued research are: the impact on student achievement of using rubrics; the features of rubrics that result in the biggest impact; and the teaching strategies for rubrics that result in the biggest impact.

---

own standards of content, clarity, practicality, and technical soundness. This includes models that illustrate strong, middle, and weak performance on each trait. Such models are not included in this paper, but are available.

[4] We understand that criteria for quality represent consensus of the larger culture. This doesn't mean however, that anything goes.

References

Arter, J., Spandel, V., and Pollard, J. (1995). *The impact of teaching students to be self-assessors of writing*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Black, P. and Wiliams, D. (1998). Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan, 80*(2), 139-148.

Borko, H., Mayfield, V., Marion, S., et al. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education, 13*, 259-278.

Clarke, D. & Stephens, M. (1996). The ripple effect: The instructional impact of the systemic introduction of performance assessment in mathematics. In M. Bierenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge*. Norwell, MA: Kluwer Academic.

Crooks, T. (1988). Impact of classroom assessment on students. *Review of Educational Research, 58*(4), 438-481.

Hillocks, G. (1986). *Research on written composition*. Fairfax, VA: National Council of Teachers of English.

Khattri, N. (1995). *Performance assessments: Observed impacts on teaching and learning. Washington DC*: Pelavin Associates.

Office of Educational Research and Improvement (1997). *Studies of educational reform: Assessment of student performance*. Washington, DC: Author.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, September, 191-198.

Figure 1

Holistic or Analytical Trait?

Product

Performance

For this product or performance, choose a scoring method:

Choose **Holistic** *OR* **Analytical Trait**

Holistic
5
4 ✓
3
2
1

Definition: One score or rating for the entire product or performance.

Examples: California Math; Riverside Math

When to Use:

- Quick snapshot of overall status or achievement
- When speed of scoring is more important than knowing how to precisely describe quality
- Simple products or performances, in which a single dimension is adequate to define quality

Disadvantages:

- Two students can get the same score for vastly different reasons
- Not as good for identifying strengths and weaknesses and planning instruction
- Not as useful for students to use

Analytical Trait	Trait 1	Trait 2	Trait 3	Trait 4
5		✓		
4	✓			
3		✓	✓	
2				
1				

Definition: Several scores or ratings for a product or performance. Each score represents an important dimension or trait of the performance or product.

Examples: *Six-Trait Model in Writing*, NAEP Special Olympics

When to Use:

- Planning instruction—shows relative strengths and weaknesses
- Teaching students the nature of a quality product or performance—they need the details
- Detailed feedback to students or parents
- When knowing how to precisely describe quality is more important than speed
- Complicated skills, products, or performances for which several dimensions are needed to be clear

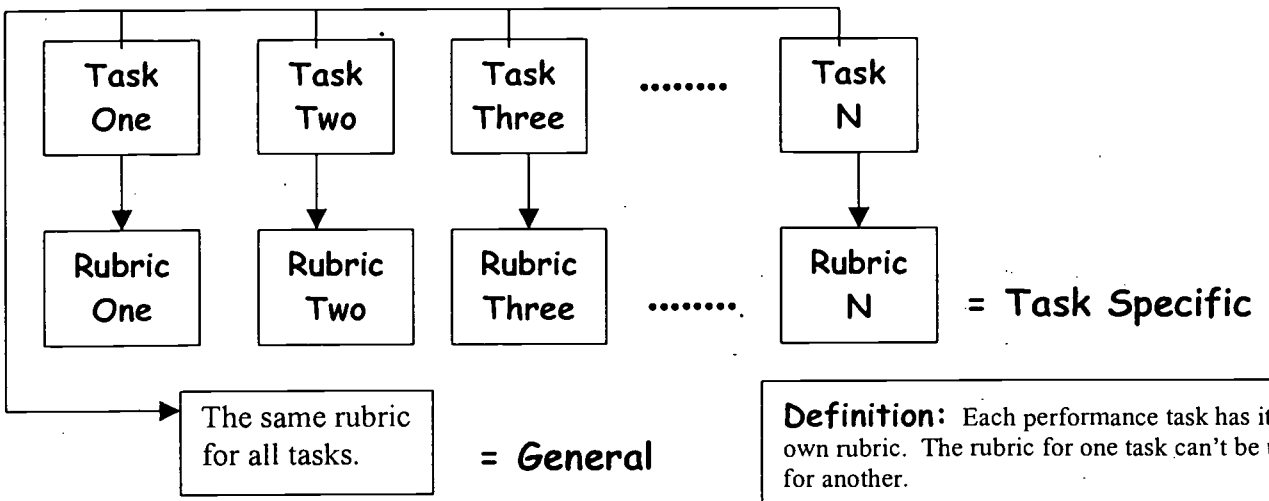
Disadvantages:

- Scoring is slower
- Takes longer to learn

Figure 2

Task-Specific or General?

For these tasks (products or performances) choose a scoring method:



Definition: Similar performance tasks use the same rubric.

Examples: California Math; Six-trait Model

When to Use:

- To define and assess reasoning, skills, and products (e.g., writing, oral presentations, critical thinking, problem solving)
- When the rubric is being used instructionally to help students understand the nature of quality and generalize from one task to the next
- When students will not all be doing exactly the same task; when students have a choice as to what evidence will be chosen to show competence on a particular skill or product
- When teachers are trying to be consistent in judging different work in different classes or grades

Disadvantages:

- Takes longer to learn; but therein is also the strength—learning the rubric is also learning the skill
- Takes longer to score, when the only goal is to assess quickly
- Sometimes task-specific features have to be added if one desires to assess knowledge of particular information

Definition: Each performance task has it's own rubric. The rubric for one task can't be used for another.

Examples: Riverside Math; NAEP Student Olympics

When to Use:

- To assess knowledge—facts, equations, methods, or procedures (e.g., describe five standards of assessment quality; use the binomial theorem to solve the following problem; describe how to use a cutting torch)
- When speed of getting a score is more important than thinking through what is being scored
- When consistency of scoring is of utmost importance

Disadvantages:

- Can't show to students ahead of time because it "gives away" the answer
- Have to develop a new rubric for each task; this takes time and sometimes this isn't even possible, e.g., portfolios
- Does not make the rater think—scoring is on automatic pilot
- "Right" answers not in the scoring guide are sometimes missed
- Does not help define the nature of quality in general; only states what quality looks like for a particular task.

California Mathematics Rubric[5]

Level 6

Solid work that may go beyond the requirements of the task(s), showing for example:

- complete understanding of the task's mathematical concepts and processes.
- clear identification of all of the important elements of the task(s).
- where appropriate, clear evidence of doing purposeful mathematics, including investigating, experimenting, modeling, designing, interpreting, analyzing, or solving.
- excellent prose and mathematical supporting arguments that may include examples or counter-examples.
- creativity and thoughtfulness in communicating the results and the interpretations of those results, to an identified audience, using dynamic and diverse means.
- multiple solutions based upon different assumptions about or interpretations of the task(s).
- unusual insights into the nature of and the resolution of problems encountered in the task(s).
- a high level of mathematical thinking that includes, where appropriate, making comparisons, conjectures, interpretations, predictions, or generalizations.
- exceptional skill in choosing appropriate mathematical tools and techniques in the resolution of problems in task(s).

Level 3

Limited completion of the requirements of the task(s), showing for example:

- an understanding of some of the task's mathematical concepts and processes, but with evidence of gaps in those understandings.
- identification of some of the important elements of the task(s), but assumptions about some of the elements may be flawed.
- communication of some ideas, but generally makes inadequate attempts to communicate, often failing to address the identified audience, and difficulty in expressing mathematical ideas.
- inadequate mathematical thinking that includes ineffective analysis procedures, limited solution strategies, unclear mathematical arguments, and inappropriate interpretation of results.
- a selection of some inappropriate tools and techniques used to resolve the situation presented in the task(s).

Level 1

Does not achieve any requirements of the task(s), showing for example:

- an irrelevant, nonsensical, or illegible response that has no valid relationship to the task(s).
- no understanding of the task's mathematical concepts and processes.
- unsuccessful attempt, if any, to communicate with the intended audience. Usually communication is not attempted.
- no attempt to explain or justify results. If attempt is made, it is often unrelated to the task, illegible, or incoherent.

---

[5] California Department of Education, (1994). *A Sampler of Mathematics Assessment-Addendum*, 721 Capital Mall, Sacramento, CA 94244. Only the definitions for Levels 6, 3 and 1 are included here for brevity sake.

IDEAS AND CONTENT[6]

(Development)

5 ***This paper is clear and focused. It holds the reader's attention. Relevant anecdotes and details enrich the central theme.***

A. The topic is **narrow** and **manageable.**
B. **Relevant, telling, quality details** give the reader important information that goes **beyond the obvious** or predictable.
C. Reasonably **accurate details** are present to support the main ideas.
D. The writer seems to be writing from **knowledge** or **experience;** the ideas are **fresh** and **original.**
E. The reader's questions are **anticipated and answered.**
F. **Insight**—an understanding of life and a knack for picking out what is significant—is an indicator of high level performance, though not required.

3 ***The writer is beginning to define the topic, even though development is still basic or general.***

A. The **topic is fairly broad;** however, you can see where the writer is headed.
B. **Support is attempted,** but doesn't go far enough yet in fleshing out the key issues or story line.
C. Ideas are **reasonably clear,** though they may not be detailed, personalized, accurate, or expanded enough to show in-depth understanding or a strong sense of purpose.
D. The writer seems to be drawing on knowledge or experience, but has **difficulty going from general observations to specifics.**
E. The reader is **left with questions.** More information is needed to "fill in the blanks."
F. The writer **generally stays on the topic** but does not develop a clear theme. The writer has not yet focused the topic past the obvious.

1 ***As yet, the paper has no clear sense of purpose or central theme. To extract meaning from the text, the reader must make inferences based on sketchy or missing details. The writing reflects more than one of these problems:***

A. The writer is **still in search of a topic,** brainstorming, or has not yet decided what the main idea of the piece will be.
B. Information is **limited** or **unclear** or the **length is not adequate** for development.
C. The idea is a **simple restatement** of the topic or an **answer** to the question with little or no attention to detail.
D. The writer has **not begun to define the topic** in a meaningful, personal way.
E. **Everything seems as important as everything else;** the reader has a hard time sifting out what is important.
F. The text may be **repetitious,** or may read like a collection of **disconnected, random thoughts** with no discernable point.

[6] From the Six Trait Model for Assessing Writing, Northwest Regional Educational Laboratory, 2000. The Six Traits are: ideas, organization, voice, word choice, sentence fluency, and conventions. Only two of the six traits are included here for brevity.

VOICE

5 ***The writer speaks directly to the reader in a way that is individual, compelling and engaging. The writer crafts the writing with an awareness and respect for the audience and the purpose for writing.***

A. The tone of the writing **adds interest** to the message and is **appropriate for the purpose and audience.**
B. The reader feels a **strong interaction** with the writer, sensing the **person behind the words.**
C. The writer **takes a risk** by revealing who he or she is consistently throughout the piece.
D. **Expository or persuasive** writing reflects a **strong commitment** to the topic by showing why the **reader needs to know this** and why he or she should care.
E. **Narrative** writing is **honest, personal, engaging** and makes you **think about and react to** the author's ideas and point of view.

3 ***The writer seems sincere, but not fully engaged or involved. The result is pleasant or even personable, but not compelling.***

A. The writer seems aware of an audience but discards personal insights in favor of **obvious generalities.**
B. The writing communicates in an **earnest, pleasing, yet safe** manner.
C. Only **one or two moments here or there** intrigue, delight, or move the reader. These places may **emerge strongly for a line or two, but quickly fade away.**
D. **Expository or persuasive** writing **lacks consistent engagement** with the topic to build credibility.
E. **Narrative** writing is **reasonably sincere,** but doesn't reflect unique or individual perspective on the topic.

1 ***The writer seems indifferent, uninvolved, or distanced from the topic and/or the audience. As a result, the paper reflects more than one of the following problems:***

A. The writer is **not concerned with the audience.** The writer's style is a **complete mismatch** for the intended reader or the writing is **so short** that little is accomplished beyond introducing the topic.
B. The writer speaks in a kind of **monotone** that flattens all potential highs or lows of the message.
C. The writing is **humdrum and "risk-free."**
D. The writing is **lifeless or mechanical;** depending on the topic, it may be overly technical or jargonistic.
E. The development of the topic is **so limited** that **no point of view is present**—zip, zero, zilch, nada.

Riverside Mathematics Example[7]

Task:

Mr. Ramirez helped four students make hey holders. He gave Rhonda, Sam, Tony, and Uta a board and told them to share it equally. First, Rhonda measured and cut one-fourth of the board. Next, Sam measured and cut one-third of the remaining board. Finally, Tony measured and cut one-half the remaining board. Uta used the piece that was left. Did the four students hare the board equally? Draw a picture and explain your answer.

Scoring Guide:

4 = Contains both a picture and an explanation that indicate a clear understanding of the pattern; contains a picture showing a whole divided into four equal parts; contains an explanation that enhances the picture by comparing the size of each piece using either sentences, computations, or a combination of both.
3 = Contains a picture that indicates a clear understanding of the pattern but only an attempt at an explanation; contains an explanation that indicates a clear understanding of the pattern, but only an attempt at a picture; has limited detail in the picture or the explanation.
2 = Offers an adequate picture only; offers an adequate explanation only; contains an explanation that does not enhance the picture; may be difficult to understand due to errors in language and grammar.
1 = Makes some attempt at a picture of an explanation; is unclear.

NAEP Special Olympics Example[8]

Task:

Joe, Sarah, Jose, Zabi, and Kim decided to hold their own Olympics after watching the Olympics on TB... They decided to have three events. They also decided to make each event equally important... The children's' scores for each of the events are listed below:

Child's Name	Frisbee Toss	Weight Lift	50-Yard Dash
Joe	40 yards	205 pounds	9.5 seconds
Jose	30 yards	170 pounds	8.0 seconds
Kim	45 yards	130 pounds	9.0 seconds
Sarah	28 yards	120 pounds	7.6 seconds
Zabi	48 yards	140 pounds	8.3 seconds

A. Who would be the overall winner?
B. Explain how you decided who would be the overall winner. Be sure to show all your work.

Rating Criteria:

4 = accurate ranking of children on each event; citing Zabi as overall winner.
3 = using a ranking approach, but misinterpreting performance on the dash event and naming incorrect winner.
2 = cites overall winner or a tie, with explanation that demonstrates recognition that a quantitative means of comparison is needed.
1 = selection of overall winner with an irrelevant, non-quantitative, or not explanation.
0 = no response.

[7] Riverside Curriculum Assessment, Performance Test Exercises, The Riverside Publishing Company, 1992.
[8] NAEP Released Exercises, 1992.

Figure 3

Build Your Own Performance Criteria

Sample Topic: Self-Reflection Letters[9]

Here's how you build generalized performance criteria. I will illustrate the steps using student self-reflection as an example. Remember, we are *not* developing these performance criteria so we can "grade" student self-reflections. We are doing it as an exercise in clear thinking about what good self-reflection looks like.

Step 1: Gather samples of student performance

Gather samples of student performance that you feel illustrate the skill or behavior in question. In our case we are trying to assess student self-reflection skills, so we might collect things that students have written in answer to questions like: "Select a piece of work that you feel illustrates your best effort. Why did you select this piece?" "What does the piece you selected show about you as a _____?" (The blank would be filled by: writer, math problem solver, life-long learner, or whatever skill students are self-reflecting about.) Other questions that might elicit student self-reflection might be: "How have you changed this year in your ability to _____?" or "What do you currently understand about _____? What do you still need to understand?"

Step 2: Sort student work into stacks; write down reasons

Place the samples of student work into three piles: strong, middle and weak. For our example we might have three piles that illustrate levels of sophistication with respect to self-reflection. As you sort the student work, write down your reasons for placing pieces in the various stacks. In other words, if you place a piece in the "sophisticated" pile, why? What cues you that the work is sophisticated? What are you saying to yourself as you place a piece of work into a pile? What might you say to a student as you return this work? These reveal criteria.

Keep sorting work until you are not adding anything new to your list of attributes. Try to create as large and diverse a list as possible. Recent lists have included such things as:

High: detailed, many things covered, insightful, self-revelation, examples provided, the student seemed motivated, sets goals for the future, looks at more than one thing, considers content as well as process, takes risks, accurate, discussion is related to criteria, growth supported with examples, organized well, sincere, honest, there are comparisons over time, shares feelings, looks at both strengths and weaknesses, there is depth to the analysis, there are good reasons and explanations, revealing, voice, easy to read, looks at skill improvement, there is ownership, there is a personal reaction, it is specific, it looks ahead, it is thorough, ideas are synthesized, it is readable, it is neat.

Middle: shows then hides, beginning of ownership, two-dimensional, some specifics, describes performance but leaves a lot out, few insights, focuses on only a few things, considers content or process but not both, is somewhat accurate, doesn't seem completely honest, there are descriptions of individual pieces of work but no comparisons over time, I have to make some inferences as to what the student meant, ...

[9] From: Regional Educational Laboratories (1998). Toolkit for Professional Developers. Portland, OR: Northwest Regional Educational Laboratory, Activity 2.1.

Low: Vague, simple restatements, mechanical, focuses mainly on surface features, obvious, same old same old, no examples, purposeless, I like it/I don't like it, one-dimensional, superficial, does not seem to be aware of the need to set goals, takes no risks, not honest, not accurate, there seems to be no ownership, I can't follow what the student is trying to say, hard to read, not organized well

Step 3: Cluster reasons into "traits" or important dimensions of performance

Cluster similar attributes together and begin to form important "traits" or "dimensions" of self-reflection. For example, the above attributes *might* be clustered into the traits of: **SKILL ANALYSIS, SINCERITY, GOAL SETTING,** and **PRESENTATION.** Number each trait and place next to each comment in the high, middle, and low categories the number of the trait to which it relates.

Make sure that your "traits" cover everything of importance. For example, **SKILL ANALYSIS** at the HIGH level might include: detailed, many things covered, insightful, examples provided, looks at more than one thing, considers content as well as process, accurate, discussion related to criteria, growth supported with examples, looks at both strengths and weaknesses, depth of analysis, revealing, specific, thorough, and synthesized ideas.

Be prepared for frequent changes, at least at the initial stages of building the criteria. Ideas need time to gel. We get new perspectives and insights as we attempt to "make sense" of our criteria.

Step 4: Write a value-neutral definition of each trait

Write a definition of each trait. For example:

> *Sincerity reflects how serious the student is about the self-reflection process. It is the degree of ownership, effort and honesty in the self-reflection. It is the degree to which the student makes a real effort to self-analyze.*

These definitions should be "value neutral"—they describe what the trait is about, not what good performance looks like. (Descriptions of good performance on the trait are left to the "high" rating.) Here is an example of the definition above converted to a statement that is *not* value-neutral: *Sincerity is taking ownership for work, really trying hard and being very honest; the student has made a sincere effort to self-reflect.*

Step 5: Find samples of student performance that illustrate each score point on each trait

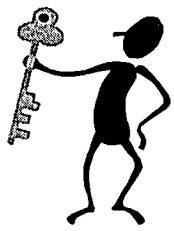
Find samples of student work which are good examples of strong, weak and mid range performance on each trait. Be sure to have several sets representing typical ranges of performance at each level of: early elementary, late elementary, middle school, and high school. These can be used to illustrate to students what to do and what "good" looks like.

Step 6: Make it better

Criteria evolve with use. Try them out. You'll probably find some parts of the criteria that work fine and some that don't. Add and modify descriptions so that they communicate better. Choose better sample papers that illustrate what you mean. Revise traits if you need to. Let students help—this is a tool for learning.

Figure 4

Criteria for Assessing Criteria (Metarubric)



This is a rubric for evaluating the quality of rubrics—a metarubric. This means that it should adhere to it's own criteria of *Content, Clarity, Practicality, and Technical Soundness.* Please note that this rubric was developed for classroom assessments, not large-scale assessments. The requirements of large-scale assessments might require features that would be counter-productive in a rubric intended for classroom use.

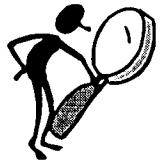
The descriptors under each trait are not meant as a checklist. Rather, they are meant as indicators that help the user focus on the correct score. An odd number of points is used because the middle score represents a balance of strengths and weaknesses—the rubric is strong in some ways, but weak in others. A strong score doesn't necessarily mean that the rubric is perfect; rather, it means that you'd have very little work to do to get it ready for your own use. A weak score means that the rubric needs so much work that it probably isn't worth the effort—it's time to find another one. It might even be easier to begin from scratch. A middle score means that the rubric is about halfway there—it would take some work to make it usable, but it probably is worth the effort.

Additionally, a middle score does not mean "average." This is a criterion-reference scale, not a norm-referenced one. It is meant to describe levels of quality in a rubric, not to describe what is currently available. It could be that the average rubric currently available is closer to a "not ready for prime time" than to a "on it's way."

The scale could easily be expanded to a five-point scale. In this case, think of "4" as a balance of characteristics from the "5" and "3" levels. Likewise, a "2" would be a balance of characteristics from the "3" and "1" levels.

No performance standard has been set on the metarubric. In other words, there is no "cut" score that indicates when a rubric is "good enough."

Trait 1: Content/Coverage



The content of a rubric defines what to look for in a student's product or performance to determine its quality. Rubric content truly is the final definition of content standards because the rubric describes what will "count"—regardless of what is stated in content standards, curriculum frameworks, or instructional materials, the content of the rubric is what teachers and students will use to determine what they need to do in order to succeed—what they see is what you'll get. Therefore, it is essential that the rubric cover all essential aspects that define quality in a product or performance and leaves out all things trivial.

Questions to ask oneself when evaluating a rubric for content are: Can I explain why each thing I have included in my rubric is essential to a quality performance? Can I cite references that describe the best thinking in the field on the nature of quality performance? Can I describe what I left out and why I left it out? Do I ever find performances or products that are scored low (or high) that I really think are good (or bad)? If so, my rubric needs to be fine-tuned to cover that which is really important.

Ready to Roll:

- There is justification for the dimensions of student performance or work that are cited as being indicators of quality. Content is based on the best thinking in the field.
- The content has the "ring of truth"—your experience as a teacher confirms that the content is truly what you *do* look for when you evaluate the quality of student work or performance.
- If counting the number of something (such as the number of references at the end of a research report) is included as an indicator, such counts really *are* indicators of quality. (Sometimes, for example, two really good references are better than 10 bad ones. Or, in writing, 10 errors in spelling all on different words is more of a problem than 10 errors in spelling all on the same word.)
- The relevant emphasis on various features of performance is right—things that are more important are stressed more; things that are less important are stressed less.
- Definitions of terms are correct—they reflect current thinking in the field.
- The number of points used in the rating scale make sense. In other words, if a five-point scale is used, is it clear why? Why not a four- or six-point scale? The level of precision is appropriate for the use.
- The developer has been selective, yet complete. There is a sense that the features of importance have been covered well, yet there is no overload.
- You are left with few questions about what was included or why it was included.
- The rubric is insightful. It really helps you organize your thinking about what it means to perform with quality. The content will help you assist students to understand the nature of quality.

On It's Way, but Needs Revision:

- The rubric is about halfway there on content. Much of the content is relevant, but you can easily think of some important things that have been left out or that have been given short shrift.
- The developer is beginning to develop the relevant aspects of performance. You can see where the rubric is headed, even though some features might not ring true or are out of balance.
- Although much of the rubric seems reasonable, some of it doesn't seem to represent current best thinking about what it means to perform well on the product or skill under consideration.
- Although the content seems fairly complete, the rubric sprawls—it's not organized very well.
- Although much of the rubric covers that which is important, it also contains several irrelevant features that might lead to an incorrect conclusion about the quality of the students' performance.

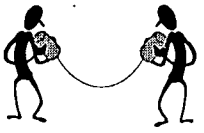
Not Ready for Prime Time:

- You can think of many important dimensions of a quality performance or product that are not included in the rubric.

- There are several irrelevant features. You find yourself asking, "Why assess this?" or "Why should this count?" or "Why is it important that students do it *this* way?"
- Content is based on counting the number of something when quality is more important than quantity.
- The rubric seems "mixed up"—things that go together don't seem to be placed together. Things that are different are put together.
- The rubric is very out of balance—features of importance are weighted incorrectly. (For example a business letter might have several categories that relate to format but only one that relates to content and organization.)
- Definitions of terms are incorrect—they don't reflect current best thinking in the field.
- The rubric is an endless list of seemingly everything the developer can think of that might be even marginally important. There is no organization to it. The developer seems unable to pick out that which is most significant or telling. The rubric looks like a brainstormed list.
- You are left with many questions about what was included and why it was included.
- There are many features of the rubric that might lead a rater to an incorrect conclusion about the quality of a student's performance.
- The rubric doesn't seem to align with the content standard it's supposed to assess.

Trait 2: Clarity

A rubric is clear to the extent that teachers, students and others are likely to interpret the statements and terms in the rubric the same way. Please notice that a rubric can be strong on the trait of content/coverage, but weak on the trait of clarity—the rubric seems to cover the important dimensions of performance, but they aren't described very well. Likewise, a rubric can be strong on the trait of clarity, but weak on the trait of content/coverage—it's very clear what the rubric means, it's just not very important stuff.



Questions to ask oneself when evaluating a rubric for clarity are: Would two teachers give the same rating on a product or performance? Can I define each statement in my rubric in such a way that students can understand what I mean? Could I find examples of student work or performances that illustrate each level of quality? Would I know what to say if a student asks, "Why did I get this score?"

Ready to Roll:

- The rubric is so clear that different teachers would give the same rating to the same product or performance.
- A single teacher could use the rubric to provide consistent ratings across assignments, time and students.
- Words are specific and accurate. It is easy to understand just what is meant.
- There are several samples of student products or performance that illustrate each score point. It is clear why each sample was scored the way it was.
- Terms are defined.

- There is just enough descriptive detail in the form of concrete indicators, adjectives, and descriptive phrases that allow you to match a student performance to the "right" score.
- There is not an overabundance of descriptive detail—the developer seems to have a sense of that which is most telling.
- The basis for assigning ratings or checkmarks is clear. Each score point is defined with indicators and descriptions.

On Its Way, but Needs Revision:

- Major headings are defined, but there is little detail to assist the rater to choose the proper score points.
- There is some attempt to define terms and include descriptors, but it doesn't go far enough.
- Teachers would agree on how to rate some things in the rubric while others are not well-defined and would probably result in disagreements.
- A single teacher would probably have trouble being consistent in scoring across students or assignments.

Not Ready For Prime Time:

- Language is so vague that almost anything could be meant. You find yourself saying things like: "I'm confused," or "I don't have any idea what they mean by this."
- There are no definitions for terms used in the rubric; or, the definitions don't help or are incorrect.
- The rubric is little more that a list of categories to rate followed by a rating scale. Nothing is defined. Few descriptors are given to define levels of performance.
- No sample student work is provided that illustrates what is meant.
- Teachers are unlikely to agree on ratings because there are so many different ways a descriptor can be interpreted.
- The only way to distinguish levels is words like: "extremely," "very," "some," "little," and "none;" or "completely," "substantially," "fairly well," "little," and "not at all."

Trait 3: Practicality

Having clear criteria that cover the right "stuff" means nothing if the system is too cumbersome to use. The trait of practicality refers to ease of use—can teachers and students understand and use it easily? Does it give them the information they need for instructional decision making and tracking student progress toward important learning outcomes? Can the rubric be used for more than just a way to assess students; can it also be used to improve the very achievement also being assessed?

Ready to Roll:

- The rubric is manageable—there are not too many things to remember; teachers and students can easily internalize them.
- It is clear how to translate results into instruction. For example, if students appear to be weak in writing, is it clear what should be taught to improve performance?
- The rubric is usually analytical trait rather than holistic, when the product or skill is complex.

- The rubric is usually general rather than task-specific. In other words, the rubric is broadly applicable to the content of interest; it is not tied to any specific exercise of assignment.
- If task-specific and/or holistic rubrics are used, their justification is clear and appropriate. Justifications could include: (a) the complexity of the skill being assessed—a "big" skill would require an analytical trait rubric while a "small skill" might only need a single "holistic" rubric; or (b) the nature of the skill being assessed—understanding a concept might require a task-specific rubric, while demonstrating a skill (such as an oral presentation) might imply a general rubric.
- The rubric can be used by students, themselves, to revise their own work, plan their own learning, and track their own progress. There is assistance on how to use the rubric in this fashion.
- There are "student-friendly" versions.
- The rubric is so clear that a student doing poorly would know exactly what to do in order to improve.
- The rubric is visually appealing to students; it draws them into its use.
- The rubric is divided into easily understandable chunks (traits) that help students grasp the essential aspects of a complex performance.
- The language used in the rubric is developmental—low scores do not imply "bad" or "failure."

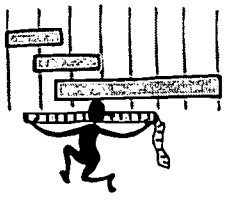
On Its Way, but Needs Revision:

- The rubric might provide useful information, but might not be easy to use.
- The rubric might be generic, yet holistic (when to be of maximal use, an analytical trait rubric would be better).
- The rubric has potential for teacher use, but would need some "tweaking"—combining long lists of attributes into traits or adjusting the language so it is clear what is intended.
- The rubric has potential for student self-use, but would need some "tweaking." This could include wording changes, streamlining, or making the format more appealing.
- Students could accurately rate their own work or performances, but it might not be clear to them what to do to improve.
- Although there are some problems, it would be easier to try and fix the rubric than look elsewhere.

Not Ready for Prime Time:

- There is no justification given for the type of rubric used—holistic or analytical trait; task-specific or generic. You get the feeling that the developer didn't know what options were available and just did what seemed like a good idea at the time.
- There seems to be no consideration of how the rubric might be useful to teachers. The intent seems to be only large-scale assessment efficiency.
- The rubric is not manageable—there is an overabundance of things to rate—it would take forever; or, everything is presented all at once and might overwhelm the user.
- It is not clear how to translate results into instruction.
- The rubric is worded in ways students would not understand.
- Fixing the rubric for student use would be harder than looking elsewhere.

Trait 4: Technical Soundness/Fairness



It is important to have "hard" evidence that the performance criteria adequately measure the goal being assessed, that it can be applied consistently, and that there is reason to believe that the ratings actually do represent what students can do. Although this might be beyond the scope of what individual classroom teachers can do, we all still have the responsibility to ask hard questions when we adopt or develop a rubric. The following are the things all educators should think about.

Ready to Roll:

- There is technical information associated with the rubric that describes rater agreement rates and the conditions under which such agreement rates can be obtained. These rater agreement rates are at least 65% exact agreement, and 98% within one point.
- The language used in the rubric is appropriate for the diversity of students found in typical classrooms. The language avoids stereotypic thinking, appeals to various learning styles, and uses language that ELL students would understand.
- There have been formal bias reviews of rubric content; studies of ratings under the various conditions in which ratings will occur, and studies, that, for example, handwriting and gender or race of the student doesn't affect judgment; etc.
- Wording is supportive of students—it describes status of a performance rather than judgements of student worth.

On Its Way, but Needs Revision:

- There is technical information associated with the rubric that describes rater agreement rates and the conditions under which such agreement rates can be obtained. These rater agreement rates aren't, however, at the levels described under "ready to roll." But, this might be due to less than adequate training of raters rather than to the scale itself.
- The language used in the rubric is inconsistent in its appropriateness for the diversity of students found in typical classrooms. But, these problems can be easily corrected.
- The authors present some hard data on the technical soundness of the rubric, but this has holes.
- Wording is inconsistently supportive of students, but could be easily corrected.

Not Ready for Prime Time:

- There is no technical information associated with the rubric.
- There have been no studies on the rubric to show that it assesses what is intended.
- The language used in the rubric is not appropriate for the diversity of students found in typical classrooms. The language might include stereotypes, appeal to some learning styles over others, and might put ELL students at a disadvantage. These problems are not easily corrected.
- The language used in the rubric might be hurtful to students. For example, at the low end of the rating scale terms such as "clueless" or "has no idea how to proceed" are used. These problems would not be easy to correct.

Figure 5

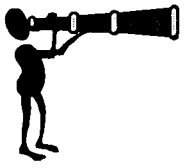
Seven Strategies for Using Criteria as a Teaching Tool[10]

1. **Teach students the vocabulary they need** to think and talk like _____ (writers, critical thinkers, mathematics problem solvers, etc.). In other words, teach students the criteria for quality. This can be accomplished by:

   a. Asking students to brainstorm characteristics of good quality work.
   b. Showing students some samples of work (of both low and high quality) and asking them to expand their list of quality features.
   c. Asking students if they'd like to see what teachers think. (They always want to.) Show them the student-friendly versions of the criteria the teacher developed and show them how what they said matches to what teachers say.

2. **Practice scoring anonymous samples of student work**. Ask students to use the rubric(s) to analyze real samples of student work. What is the author of this work doing well and what could be improved? Remember, there is no single correct score, there is only a justifiable score. Ask students to justify their scores using wording from the rubric. Begin with single traits. When students are proficient at single traits ask them to analyze work for more than one trait at a time.

3. **Practice focused revision**. It's not enough to merely ask students to judge work and justify their judgments. Students also need to understand how to improve work that is weak. Ask students to give the author advice on how to improve his or her work on the trait under consideration. Brainstorm a list on the board. Then ask students to work in groups to revise the work using their own advice. Again, begin with single traits; pick work that needs revision on a particular trait. Revisions will almost always be better on more than just the single trait being addressed. For example, as students improve ideas (in writing), organization and voice tend to improve also.



4. **Share professional examples.** Show students lots of samples from the real world—both strong and weak. Have them analyze these samples for quality.

5. **Model it yourself.** Ask students to analyze your work for quality and make suggestions on improvement. Take the students' advice and revise your own work. Ask them to again review it for quality. Students *love* doing this.

[10] Based on work done at Northwest Regional Educational Laboratory, Portland, OR, and Jan Chappuis, assessment consultant, Silverdale, WA.

6. **Give students plenty of opportunities to show what they know**. We all learn best when we can describe and articulate the criteria for quality—ask them to develop posters or drawings that illustrate the traits of quality; write letter to their parents describing what they have done well and their next steps (using the language of the criteria); etc.; ask students to reflect on their work using the language of the criteria.

7. **Focus lessons on the traits of quality**. Use the criteria and traits as organizers for lessons a. Make sure that every lesson explicitly addresses one or more traits. For example, in writing you could teach specific lessons on organization—different organizational patterns and when to use them; transitions; catchy openings; good endings; etc. Make sure that students understand that this lesson will help them to organize their writing better. You already teach these lessons. The main difference is being very clear which trait(s) each lesson addresses.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM031843

I. DOCUMENT IDENTIFICATION:

Title: Rubrics, Scoring Guides and Performance Criteria: Classroom Tools for Assessing and Improving Student Learning

Author(s): Judith A. Arter

Corporate Source: Assessment Training Institute, Portland, OR

Publication Date: 4/24/2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Judith A. Arter

Printed Name/Position/Title: Judith A. Arter

Organization/Address: ATI, 50 SW 2nd Ave, Suite 300, Portland, OR 97204

Telephone: 503-228-3060 FAX: 503-228-3014

E-Mail Address: judya@assessmentinst.com Date: 9/12/00

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)
PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE